

平成22年 6月 9日現在

研究種目：若手研究（スタートアップ）

研究期間：2008～2009

課題番号：20800013

研究課題名（和文）医学論文データベースからの知識ネットワーク抽出に関する研究

研究課題名（英文）Extraction of Knowledge Network from Medical Publication Database

研究代表者

岡 瑞起（MIZUKI OKA）

東京大学・知の構造化センター・特任研究員

研究者番号：10512105

研究成果の概要（和文）：本研究では、論文情報のオンラインデータ化や公開アクセスが十分行われていない分野の一つである医学論文を対象に、データからの情報抽出、抽出された要素関係の分析を行なう手法の開発を行った。開発した技術が、今後実用化されると、分野の専門家だけでなくもあるトピックにおける関連研究者は誰であるのかを簡単に調査できたり、重要論文を見つけられたりといった学術に関わる調査をサポートすることがより容易なることが期待される。

研究成果の概要（英文）：We have developed technologies for extracting named-entities, their relations from medical publication data. When these technologies are put in place for real usage, it will be able to support finding important elements such as researchers and papers in a particular field.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
20年度	1,200,000	360,000	1,560,000
21年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総計	2,400,000	720,000	3,120,000

研究分野：

科研費の分科・細目：

キーワード：学術情報，関係抽出，可視化

1. 研究開始当初の背景

WWW(Web)の普及により、様々なデータが電子化され公開されることにより誰にでもアクセス可能な状況になってきている。そのような流れの一つとして、学術論文に関する情報のデジタルオンライン化が進められている。特に、コンピュータサイエンスの世界では、そのようなデジタル化が精力的に行

われ情報が公開されており、研究者間や会議間の関係抽出、また研究トピック抽出の研究が国内外を問わず活発に行われている。一方、まだそのような論文情報のオンラインデータ化や公開アクセスが十分行われていない分野も多数ある。データがデジタル化され簡易アクセスが可能になれば、重要研究者、会議、トピックの抽出といった情報処理が実現

可能であり、新しい事実発見に寄与でき、また学術知識の効率的な共有が可能になる。そのようなデータ化が十分に行われていない分野として医学論文が挙げられる。

2. 研究の目的

このような状況から、本研究では医学論文を対象とし、それらの論文から情報（著者、文献情報、トピック、雑誌／会議名、アブストラクト、参考文献リスト）を抽出し、これら複数の要素の関係を解析することにより、多岐の分野に渡る医学論文の俯瞰図を与えることを目標とする。

3. 研究の方法

本提案の目的は、専門家・細分化が顕著な医学論文データから、著者、タイトル、所属名、論文誌／会議名、トピック、文献リスト等の要素を抽出し、それらの情報を構造化することである。そのために、まず、医学論文データベースを作成し、次に要素間の関係を抽出する手法、また、医学における学術知識の俯瞰図を与える手法を研究／開発し、実験／考察を行った。

4. 研究成果

ウェブの普及により、様々なデータが電子化され公開されることにより誰にでもアクセス可能な状況になってきている。そのような流れの一つとして、学術論文に関する情報のデジタルオンライン化が進められている。特に、コンピュータサイエンスの世界では、そのようなデジタル化が精力的に行われ情報が公開されてきており、研究者間や会議間の関係抽出、また研究トピック抽出の研究が国内外を問わず活発に行われている。一方、そのような論文情報のオンラインデータ化や公開アクセスが十分行われていない分野の一つに医学論文が挙げられる。学術情報からの情報抽出、抽出された要素関係の分析は、コンピュータサイエンスの論文文献データにおいて盛んに扱われているが、本研究の対象とする医学論文データはその性質が細分化・専門化されているという点にて大きく異なる。本研究では、このようなデータからの固有表現の抽出とそれらの間の関係抽出の新しい手法の開発、および、抽出された情報を可視化する技術の開発を行った。開発した技術を応用することにより、分野の専門家でなくてもあるトピックにおける関連研究者は誰であるのかを簡単に調査できたり、重要論文を見つけられたりといった学術に関わる調査をサポートすることがより容易になることが期待される。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者に

は下線)

[雑誌論文] (計 1 件)

①岡瑞起, 松尾豊, 検索エンジンを用いた関係の重み付け, 人工知能学会論文誌, No. 25, pp. 1-8, 2010

[学会発表] (計 5 件)

①岡瑞起, 李明喜, 橋本康弘, 江渡浩一郎, 環境デザインにおけるパターン・ランゲージ, The First Annual Asian PLoP, 2010

②荒牧英治, 橋本康弘, 宇野良子, 河村美雪, 石橋素, 李明喜, 岡瑞起, pingpong:新しいデザインのための行為抽出エンジン, NLP 若手の会第 4 回シンポジウム, 2009

③荒牧英治, 宇野良子, 岡瑞起, TYPO Writer: ヒトはどのように打ち間違えるのか?, 言語処理学会第 16 回年次大会, 2010

④Mizuki Oka, Yutaka Matsuo, Weighting Relations in Social Networks Using the Web, 人工知能学会全国大会, 2009

⑤Mizuki Oka, Yutaka Matsuo, Measuring the Weight of Relations Between Entities, Third International Workshop on Service Matchmaking and Resource Retrieval in the Semantic Web, 2009

6. 研究組織

(1) 研究代表者

岡 瑞起 (MIZUKI OKA)

東京大学・知の構造化センター・特任研究員

研究者番号: 10512105